Characterizations of finite and infinite episturmian words via lexicographic orderings

Amy Glen*, Jacques Justin, Giuseppe Pirillo

LaCIM Seminar – Friday 8 December, 2006

Outline

Episturmian words

- An interesting natural generalization of the well-known *Sturmian words*.
- Share many properties with Sturmian words.
- Include the well-known *Arnoux-Rauzy sequences*.
- Introduced by X. Droubay, J. Justin, and G. Pirillo (2001).

Main results

We characterize by *lexicographic order* all:

- *finite* Sturmian and episturmian words;
- episturmian words in a *wide sense* (recurrent, episkew);
- *balanced* infinite words over a 2-letter alphabet.

Outline

Words

Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1 x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- *Length*: $|u| = m$
- $|u|_a$: number of occurrences of the letter a in u
- *Reversal*: $\widetilde{u} = x_m x_{m-1} \cdots x_1$
- u is a *palindrome* if $u = \widetilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- ε: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$

Words

Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1x_2\cdots x_m$, each $x_i \in \mathcal{A}$.

- *Length*: $|u| = m$
- $|u|_a$: number of occurrences of the letter a in u
- *Reversal*: $\widetilde{u} = x_mx_{m-1}\cdots x_1$
- u is a *palindrome* if $u = \widetilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- ε: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$

Words

Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1x_2\cdots x_m$, each $x_i \in \mathcal{A}$.

- *Length*: $|u| = m$
- $|u|_a$: number of occurrences of the letter a in u
- *Reversal*: $\widetilde{u} = x_mx_{m-1}\cdots x_1$
- u is a *palindrome* if $u = \widetilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- ε: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$

Words

Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1x_2\cdots x_m$, each $x_i \in \mathcal{A}$.

- *Length*: $|u| = m$
- $|u|_a$: number of occurrences of the letter a in u
- *Reversal*: $\widetilde{u} = x_m x_{m-1} \cdots x_1$
- u is a *palindrome* if $u = \widetilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- ε: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$

Words

Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1 x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- *Length*: $|u| = m$
- $|u|_a$: number of occurrences of the letter a in u
- *Reversal*: $\widetilde{u} = x_m x_{m-1} \cdots x_1$
- u is a *palindrome* if $u = \widetilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- ε: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$

Words

Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1x_2\cdots x_m$, each $x_i \in \mathcal{A}$.

- *Length*: $|u| = m$
- $|u|_a$: number of occurrences of the letter a in u
- *Reversal*: $\widetilde{u} = x_mx_{m-1}\cdots x_1$
- u is a *palindrome* if $u = \widetilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- ε: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$

Words

Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1 x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- *Length*: $|u| = m$
- $|u|_a$: number of occurrences of the letter a in u
- *Reversal*: $\widetilde{u} = x_m x_{m-1} \cdots x_1$
- u is a *palindrome* if $u = \widetilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- ε: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$

Words

Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1x_2\cdots x_m$, each $x_i \in \mathcal{A}$.

- *Length*: $|u| = m$
- $|u|_a$: number of occurrences of the letter a in u
- *Reversal*: $\widetilde{u} = x_mx_{m-1}\cdots x_1$
- u is a *palindrome* if $u = \widetilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- ε: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$

Words (cont.)

Let $\mathbf{x} = x_0x_1x_2\cdots$ be an *infinite word* over $\mathcal{A}$.

- *Factor of* $\mathbf{x}$: a finite string of consecutive letters in $\mathbf{x}$
- *Prefix of* $\mathbf{x}$: factor occurring at the beginning of $\mathbf{x}$
- $F(\mathbf{x})$: *set of all factors* of $\mathbf{x}$
- $\mathrm{Ult}(\mathbf{x})$: set of letters occurring infinitely often in $\mathbf{x}$
- $\mathrm{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is *recurrent* in $\mathbf{x}$ if w occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is *recurrent* if all of its factors are recurrent in it

Words (cont.)

Let $\mathbf{x} = x_0x_1x_2\cdots$ be an *infinite word* over $\mathcal{A}$.

- *Factor of* $\mathbf{x}$: a finite string of consecutive letters in $\mathbf{x}$
- *Prefix of* $\mathbf{x}$: factor occurring at the beginning of $\mathbf{x}$
- $F(\mathbf{x})$: *set of all factors* of $\mathbf{x}$
- $\mathrm{Ult}(\mathbf{x})$: set of letters occurring infinitely often in $\mathbf{x}$
- $\mathrm{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is *recurrent* in $\mathbf{x}$ if w occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is *recurrent* if all of its factors are recurrent in it

Words (cont.)

Let $\mathbf{x} = x_0 x_1 x_2 \cdots$ be an *infinite word* over $\mathcal{A}$.

- *Factor of* $\mathbf{x}$: a finite string of consecutive letters in $\mathbf{x}$
- *Prefix of* $\mathbf{x}$: factor occurring at the beginning of $\mathbf{x}$
- $F(\mathbf{x})$: *set of all factors* of $\mathbf{x}$
- $\mathrm{Ult}(\mathbf{x})$: set of letters occurring infinitely often in $\mathbf{x}$
- $\mathrm{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is *recurrent* in $\mathbf{x}$ if w occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is *recurrent* if all of its factors are recurrent in it

Words (cont.)

Let $\mathbf{x} = x_0x_1x_2\cdots$ be an *infinite word* over $\mathcal{A}$.

- *Factor of* $\mathbf{x}$: a finite string of consecutive letters in $\mathbf{x}$
- *Prefix of* $\mathbf{x}$: factor occurring at the beginning of $\mathbf{x}$
- $F(\mathbf{x})$: *set of all factors* of $\mathbf{x}$
- $\mathrm{Ult}(\mathbf{x})$: set of letters occurring infinitely often in $\mathbf{x}$
- $\mathrm{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is *recurrent* in $\mathbf{x}$ if w occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is *recurrent* if all of its factors are recurrent in it

Words (cont.)

Let $\mathbf{x} = x_0 x_1 x_2 \cdots$ be an *infinite word* over $\mathcal{A}$.

- *Factor of* $\mathbf{x}$: a finite string of consecutive letters in $\mathbf{x}$
- *Prefix of* $\mathbf{x}$: factor occurring at the beginning of $\mathbf{x}$
- $F(\mathbf{x})$: *set of all factors* of $\mathbf{x}$
- $\mathrm{Ult}(\mathbf{x})$: set of letters occurring infinitely often in $\mathbf{x}$
- $\mathrm{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is *recurrent* in $\mathbf{x}$ if w occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is *recurrent* if all of its factors are recurrent in it

Words (cont.)

Let $\mathbf{x} = x_0x_1x_2\cdots$ be an *infinite word* over $\mathcal{A}$.

- *Factor of* $\mathbf{x}$: a finite string of consecutive letters in $\mathbf{x}$
- *Prefix of* $\mathbf{x}$: factor occurring at the beginning of $\mathbf{x}$
- $F(\mathbf{x})$: *set of all factors* of $\mathbf{x}$
- $\mathrm{Ult}(\mathbf{x})$: set of letters occurring infinitely often in $\mathbf{x}$
- $\mathrm{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is *recurrent* in $\mathbf{x}$ if w occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is *recurrent* if all of its factors are recurrent in it

Words (cont.)

Let $\mathbf{x} = x_0 x_1 x_2 \cdots$ be an *infinite word* over $\mathcal{A}$.

- *Factor of* $\mathbf{x}$: a finite string of consecutive letters in $\mathbf{x}$
- *Prefix of* $\mathbf{x}$: factor occurring at the beginning of $\mathbf{x}$
- $F(\mathbf{x})$: *set of all factors* of $\mathbf{x}$
- $\mathrm{Ult}(\mathbf{x})$: set of letters occurring infinitely often in $\mathbf{x}$
- $\mathrm{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is *recurrent* in $\mathbf{x}$ if w occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is *recurrent* if all of its factors are recurrent in it

Words (cont.)

Let $\mathbf{x} = x_0x_1x_2\cdots$ be an *infinite word* over $\mathcal{A}$.

- *Factor of* $\mathbf{x}$: a finite string of consecutive letters in $\mathbf{x}$
- *Prefix of* $\mathbf{x}$: factor occurring at the beginning of $\mathbf{x}$
- $F(\mathbf{x})$: *set of all factors* of $\mathbf{x}$
- $\mathrm{Ult}(\mathbf{x})$: set of letters occurring infinitely often in $\mathbf{x}$
- $\mathrm{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is *recurrent* in $\mathbf{x}$ if w occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is *recurrent* if all of its factors are recurrent in it

Lexicographic Order

Suppose $\mathcal{A}$ is totally ordered by the relation $<$. Then we can totally order $\mathcal{A}^+$ by the *lexicographic order* $<$.

That is:

Definition

Given two words $u, v \in \mathcal{A}^+$, we have $u < v \Leftrightarrow$ either u is a proper prefix of v or $u = xau'$ and $v = xbv'$, for some x, u', $v' \in \mathcal{A}^*$ and letters a, b with $a < b$.

- This is the usual alphabetic ordering in a dictionary.
- We say that u is *lexicographically less* than v.
- This notion naturally extends to infinite words.

Lexicographic Order

Suppose $\mathcal{A}$ is totally ordered by the relation $<$. Then we can totally order $\mathcal{A}^+$ by the *lexicographic order* $<$.

That is:

Definition

Given two words $u, v \in \mathcal{A}^+$, we have $u < v \Leftrightarrow$ either u is a proper prefix of v or $u = xau'$ and $v = xbv'$, for some x, u', $v' \in \mathcal{A}^*$ and letters a, b with $a < b$.

- This is the usual alphabetic ordering in a dictionary.
- We say that u is *lexicographically less* than v.
- This notion naturally extends to infinite words.

Lexicographic Order

Suppose $\mathcal{A}$ is totally ordered by the relation $<$. Then we can totally order $\mathcal{A}^+$ by the *lexicographic order* $<$.

That is:

Definition

Given two words $u, v \in \mathcal{A}^+$, we have $u < v \Leftrightarrow$ either u is a proper prefix of v or $u = xau'$ and $v = xbv'$, for some x, u', $v' \in \mathcal{A}^*$ and letters a, b with $a < b$.

- This is the usual alphabetic ordering in a dictionary.
- We say that u is *lexicographically less* than v.
- This notion naturally extends to infinite words.

Outline

Sturmian words

Definition

An infinite word **s** over $\{a, b\}$ is *Sturmian* if there exist real numbers α, $\rho \in [0, 1]$ such that **s** is equal to one of the following two infinite words:

$$s_{\alpha,\rho},\ s'_{\alpha,\rho} : \mathbb{N} \rightarrow \{a, b\}$$

defined by

$$\left.\begin{array}{l} s_{\alpha,\rho}(n) = \begin{cases} a & \text{if } \lfloor (n+1)\alpha + \rho \rfloor - \lfloor n\alpha + \rho \rfloor = 0, \\ b & \text{otherwise;} \end{cases} \\ \\ s'_{\alpha,\rho}(n) = \begin{cases} a & \text{if } \lceil (n+1)\alpha + \rho \rceil - \lceil n\alpha + \rho \rceil = 0, \\ b & \text{otherwise.} \end{cases} \end{array}\right. \qquad (n \geq 0)$$

Sturmian words (cont.)

A *Sturmian word* is:

- *aperiodic* if α is irrational;
- *periodic* if α is rational;
- *standard* if $\rho = \alpha$.

Here, *Sturmian* refers to both aperiodic and periodic Sturmian words.

Sturmian words (cont.)

A *Sturmian word* is:

- *aperiodic* if α is irrational;
- *periodic* if α is rational;
- *standard* if $\rho = \alpha$.

Here, *Sturmian* refers to both aperiodic and periodic Sturmian words.

Sturmian words (cont.)

Definition (Balance)

A finite or infinite word w on $\{a, b\}$ is *balanced* if:

$$u,\ v \in F(w),\ |u| = |v| \quad \Rightarrow \quad ||u|_b - |v|_b| \leq 1.$$

Morse & Hedlund (1940):

All balanced infinite words over a 2-letter alphabet are called *Sturmian trajectories*.

They belong to three classes:

- aperiodic Sturmian;
- periodic Sturmian;
- ultimately periodic non-recurrent infinite words, called *skew words*.

Sturmian words (cont.)

Definition (Balance)

A finite or infinite word w on $\{a, b\}$ is *balanced* if:

$$u,\ v \in F(w),\ |u| = |v| \quad \Rightarrow \quad ||u|_b - |v|_b| \leq 1.$$

Morse & Hedlund (1940):

All balanced infinite words over a 2-letter alphabet are called *Sturmian trajectories*.

They belong to three classes:

- aperiodic Sturmian;
- periodic Sturmian;
- ultimately periodic non-recurrent infinite words, called *skew words*.

Episturmian words

Definition

An infinite word $\mathbf{t}$ is *episturmian* if:

- $F(\mathbf{t})$ is *closed under reversal*, and
- $\mathbf{t}$ has at most one *right special factor* of each length.

$\mathbf{t}$ is *standard* if all of its left special factors are prefixes of it.

- Episturmian words are recurrent.

Episturmian words

Definition

An infinite word $\mathbf{t}$ is *episturmian* if:

- $F(\mathbf{t})$ is *closed under reversal*, and
- $\mathbf{t}$ has at most one *right special factor* of each length.

$\mathbf{t}$ is *standard* if all of its left special factors are prefixes of it.

- Episturmian words are recurrent.

Standard episturmian words

- Let $\mathbf{t}$ be a standard episturmian word over $\mathcal{A}$ and let

$$u_1 = \varepsilon, \ u_2, \ u_3, \ u_4, \ \ldots$$

be the infinite sequence of its palindromic prefixes.

- $\exists$ an infinite word $\Delta(\mathbf{t}) = x_1 x_2 x_3 \cdots \ (x_i \in \mathcal{A})$ such that

$$\boxed{u_{n+1} = (u_n x_n)^{(+)}, \quad n \in \mathbb{N}^+}$$

where $w^{(+)}$ is the shortest palindrome having w as a prefix.

- $\Delta(\mathbf{t})$ is called the *directive word* of $\mathbf{t} = \lim_{n\to\infty} u_n$.

Standard episturmian words

- Let $\mathbf{t}$ be a standard episturmian word over $\mathcal{A}$ and let

$$u_1 = \varepsilon, \ u_2, \ u_3, \ u_4, \ \ldots$$

be the infinite sequence of its palindromic prefixes.

- $\exists$ an infinite word $\Delta(\mathbf{t}) = x_1x_2x_3\cdots$ $(x_i \in \mathcal{A})$ such that

$$\boxed{u_{n+1} = (u_nx_n)^{(+)}, \quad n \in \mathbb{N}^+}$$

where $w^{(+)}$ is the shortest palindrome having w as a prefix.

- $\Delta(\mathbf{t})$ is called the *directive word* of $\mathbf{t} = \lim_{n\to\infty} u_n$.

Standard episturmian words

- Let $\mathbf{t}$ be a standard episturmian word over $\mathcal{A}$ and let

$$u_1 = \varepsilon, \; u_2, \; u_3, \; u_4, \; \ldots$$

  be the infinite sequence of its palindromic prefixes.

- $\exists$ an infinite word $\Delta(\mathbf{t}) = x_1 x_2 x_3 \cdots$ $(x_i \in \mathcal{A})$ such that

$$\boxed{u_{n+1} = (u_n x_n)^{(+)}, \quad n \in \mathbb{N}^+}$$

  where $w^{(+)}$ is the shortest palindrome having w as a prefix.

- $\Delta(\mathbf{t})$ is called the *directive word* of $\mathbf{t} = \lim_{n\to\infty} u_n$.

Strict episturmian words

Definition

A standard episturmian word $\mathbf{t}$ over $\mathcal{A}$, or any equivalent (episturmian) word, is said to be *$\mathcal{B}$-strict* (or *k-strict* if $|\mathcal{B}| = k$) if

$$\boxed{\mathrm{Alph}(\Delta(\mathbf{t})) = \mathrm{Ult}(\Delta(\mathbf{t})) = \mathcal{B} \subseteq \mathcal{A}}$$

- The k-strict episturmian words have *complexity* $(k-1)n+1$ for each $n \in \mathbb{N}$.
- Such words are exactly the k-letter *Arnoux-Rauzy sequences*.
- Example: *k-bonacci word*, $k \geq 2$.

Strict episturmian words

Definition

A standard episturmian word $\mathbf{t}$ over $\mathcal{A}$, or any equivalent (episturmian) word, is said to be *$\mathcal{B}$-strict* (or *k-strict* if $|\mathcal{B}| = k$) if

$$\boxed{\mathrm{Alph}(\Delta(\mathbf{t})) = \mathrm{Ult}(\Delta(\mathbf{t})) = \mathcal{B} \subseteq \mathcal{A}}$$

- The k-strict episturmian words have *complexity* $(k-1)n + 1$ for each $n \in \mathbb{N}$.
- Such words are exactly the k-letter *Arnoux-Rauzy sequences*.
- Example: *k-bonacci word*, $k \geq 2$.

Strict episturmian words

Definition

A standard episturmian word $\mathbf{t}$ over $\mathcal{A}$, or any equivalent (episturmian) word, is said to be *$\mathcal{B}$-strict* (or *k-strict* if $|\mathcal{B}| = k$) if

$$\boxed{\mathrm{Alph}(\Delta(\mathbf{t})) = \mathrm{Ult}(\Delta(\mathbf{t})) = \mathcal{B} \subseteq \mathcal{A}}$$

- The k-strict episturmian words have *complexity* $(k-1)n + 1$ for each $n \in \mathbb{N}$.
- Such words are exactly the k-letter *Arnoux-Rauzy sequences*.
- Example: *k-bonacci word*, $k \geq 2$.

Strict episturmian words

Definition

A standard episturmian word $\mathbf{t}$ over $\mathcal{A}$, or any equivalent (episturmian) word, is said to be *$\mathcal{B}$-strict* (or *k-strict* if $|\mathcal{B}| = k$) if

$$\boxed{\mathrm{Alph}(\Delta(\mathbf{t})) = \mathrm{Ult}(\Delta(\mathbf{t})) = \mathcal{B} \subseteq \mathcal{A}}$$

- The k-strict episturmian words have *complexity* $(k-1)n + 1$ for each $n \in \mathbb{N}$.
- Such words are exactly the k-letter *Arnoux-Rauzy sequences*.
- Example: *k-bonacci word*, $k \geq 2$.

Outline

Terminology

Definition

A finite word w is said to be *finite Sturmian* or *finite episturmian* if w is a factor of some infinite Sturmian or episturmian word.

- It suffices to consider strict standard episturmian words.
- Finite episturmian words are exactly the *finite Arnoux-Rauzy words* (Mignosi and Zamboni, 2002).

Definition (Episkew)

An infinite word $\mathbf{t}$ on a finite alphabet is said to be *episkew* if $\mathbf{t}$ is non-recurrent and all of its factors are (finite) episturmian.

- There are a number of equivalent definitions of episkew words.

Terminology

Definition

A finite word w is said to be *finite Sturmian* or *finite episturmian* if w is a factor of some infinite Sturmian or episturmian word.

- It suffices to consider strict standard episturmian words.
- Finite episturmian words are exactly the *finite Arnoux-Rauzy words* (Mignosi and Zamboni, 2002).

Definition (Episkew)

An infinite word $\mathbf{t}$ on a finite alphabet is said to be *episkew* if $\mathbf{t}$ is non-recurrent and all of its factors are (finite) episturmian.

- There are a number of equivalent definitions of episkew words.

Terminology

Definition

A finite word w is said to be *finite Sturmian* or *finite episturmian* if w is a factor of some infinite Sturmian or episturmian word.

- It suffices to consider strict standard episturmian words.
- Finite episturmian words are exactly the *finite Arnoux-Rauzy words* (Mignosi and Zamboni, 2002).

Definition (Episkew)

An infinite word $\mathbf{t}$ on a finite alphabet is said to be episkew if $\mathbf{t}$ is non-recurrent and all of its factors are (finite) episturmian.

- There are a number of equivalent definitions of episkew words.

Terminology

Definition

A finite word w is said to be *finite Sturmian* or *finite episturmian* if w is a factor of some infinite Sturmian or episturmian word.

- It suffices to consider strict standard episturmian words.
- Finite episturmian words are exactly the *finite Arnoux-Rauzy words* (Mignosi and Zamboni, 2002).

Definition (Episkew)

An infinite word $\mathbf{t}$ on a finite alphabet is said to be episkew if $\mathbf{t}$ is non-recurrent and all of its factors are (finite) episturmian.

- There are a number of equivalent definitions of episkew words.

Outline

1. Introduction
2. Preliminaries
   - Terminology & Notation
   - Sturmian & Episturmian Words
   - Episkew Words
3. Previous Work
   - Extremal Words
   - Extremal Properties
   - Fine Words
4. Characterizations via Lexicographic Orderings
   - Finite Episturmian Words
   - Infinite Episturmian Words

Extremal words

- Let $\mathbf{t}$ be an infinite word.

Definition

Define $\min(\mathbf{t})$ to be the infinite word such that any prefix of $\min(\mathbf{t})$ is the *lexicographically* smallest amongst the factors of $\mathbf{t}$ of the same length. Similarly define $\max(\mathbf{t})$.

- Our main results extend the following recent work . . .

Extremal words

- Let $\mathbf{t}$ be an infinite word.

Definition

Define $\min(\mathbf{t})$ to be the infinite word such that any prefix of $\min(\mathbf{t})$ is the *lexicographically* smallest amongst the factors of $\mathbf{t}$ of the same length. Similarly define $\max(\mathbf{t})$.

- Our main results extend the following recent work . . .

Outline

Extremal properties

Proposition (Pirillo, 2005)

Let $\mathbf{s}$ be an infinite word over a finite alphabet $\mathcal{A}$.
The following properties are equivalent:

(i) $\mathbf{s}$ *is standard episturmian,*

(ii) *for any $a \in \mathcal{A}$ and order $<$ such that $a = \min(\mathcal{A})$, we have $a\mathbf{s} \le \min(\mathbf{s})$.*

- Similarly, for an infinite word $\mathbf{s}$ on $\{a, b\}$ $(a < b)$, the inequality:

$$a\mathbf{s} \le \min(\mathbf{s}) \le \max(\mathbf{s}) \le b\mathbf{s}$$

characterizes *standard Sturmian words* (*aperiodic* and *periodic*).

Extremal properties

Proposition (Pirillo, 2005)

Let $\mathbf{s}$ be an infinite word over a finite alphabet $\mathcal{A}$. The following properties are equivalent:

(i) $\mathbf{s}$ *is standard episturmian,*

(ii) *for any $a \in \mathcal{A}$ and order $<$ such that $a = \min(\mathcal{A})$, we have $a\mathbf{s} \leq \min(\mathbf{s})$.*

- Similarly, for an infinite word $\mathbf{s}$ on $\{a, b\}$ ($a < b$), the inequality:

$$a\mathbf{s} \leq \min(\mathbf{s}) \leq \max(\mathbf{s}) \leq b\mathbf{s}$$

characterizes *standard Sturmian words* (*aperiodic* and *periodic*).

Proposition (Justin & Pirillo, 2002)

Let $\mathbf{s}$ *be an infinite word over a finite alphabet* $\mathcal{A}$*.*
The following properties are equivalent:

(i) $\mathbf{s}$ *is a standard Arnoux-Rauzy sequence,*

(ii) *for any* $a \in \mathcal{A}$ *and order* $<$ *such that* $a = \min(\mathcal{A})$*, we have* $a\mathbf{s} = \min(\mathbf{s})$*.*

- That is: $\mathbf{s}$ is a strict standard episturmian word $\Leftrightarrow$ (ii) holds.
- 2-letters: $\mathbf{s}$ is an aperodic standard Sturmian word $\Leftrightarrow$ $(\min(\mathbf{s}), \max(\mathbf{s})) = (a\mathbf{s}, b\mathbf{s})$ for $a < b$.

Proposition (Justin & Pirillo, 2002)

Let $\mathbf{s}$ *be an infinite word over a finite alphabet* $\mathcal{A}$.
The following properties are equivalent:

(i) $\mathbf{s}$ *is a standard Arnoux-Rauzy sequence,*

(ii) *for any* $a \in \mathcal{A}$ *and order* $<$ *such that* $a = \min(\mathcal{A})$, *we have* $a\mathbf{s} = \min(\mathbf{s})$.

- That is: $\mathbf{s}$ is a strict standard episturmian word $\Leftrightarrow$ (ii) holds.
- 2-letters: $\mathbf{s}$ is an aperodic standard Sturmian word $\Leftrightarrow$ $(\min(\mathbf{s}), \max(\mathbf{s})) = (a\mathbf{s}, b\mathbf{s})$ for $a < b$.

Proposition (Justin & Pirillo, 2002)

Let $\mathbf{s}$ *be an infinite word over a finite alphabet* $\mathcal{A}$*.*
The following properties are equivalent:

(i) $\mathbf{s}$ *is a standard Arnoux-Rauzy sequence,*

(ii) *for any* $a \in \mathcal{A}$ *and order* $<$ *such that* $a = \min(\mathcal{A})$*, we have* $a\mathbf{s} = \min(\mathbf{s})$*.*

- That is: $\mathbf{s}$ is a strict standard episturmian word $\Leftrightarrow$ (ii) holds.
- 2-letters: $\mathbf{s}$ is an aperodic standard Sturmian word $\Leftrightarrow$ $(\min(\mathbf{s}), \max(\mathbf{s})) = (a\mathbf{s}, b\mathbf{s})$ for $a < b$.

Outline

Fine words

Definition (Pirillo, 2005)

An infinite word $\mathbf{t}$ over a 2-letter alphabet $\{a, b\}$ ($a < b$) is *fine* if $(\min(\mathbf{t}), \max(\mathbf{t})) = (a\mathbf{s}, b\mathbf{s})$ for some infinite word $\mathbf{s}$.

- Fine words on $\{a, b\}$ are exactly the aperiodic Sturmian and skew infinite words.
- Recently generalized to an arbitrary finite alphabet . . .

Fine words

Definition (Pirillo, 2005)

An infinite word $\mathbf{t}$ over a 2-letter alphabet $\{a, b\}$ $(a < b)$ is *fine* if $(\min(\mathbf{t}), \max(\mathbf{t})) = (a\mathbf{s}, b\mathbf{s})$ for some infinite word $\mathbf{s}$.

- Fine words on $\{a, b\}$ are exactly the aperiodic Sturmian and skew infinite words.
- Recently generalized to an arbitrary finite alphabet . . .

Fine words

Definition (Pirillo, 2005)

An infinite word $\mathbf{t}$ over a 2-letter alphabet $\{a, b\}$ $(a < b)$ is *fine* if $(\min(\mathbf{t}), \max(\mathbf{t})) = (a\mathbf{s}, b\mathbf{s})$ for some infinite word $\mathbf{s}$.

- Fine words on $\{a, b\}$ are exactly the aperiodic Sturmian and skew infinite words.
- Recently generalized to an arbitrary finite alphabet ...

Generalized fine words

Definition

An *acceptable pair* is a pair $(a, <)$ where a is a letter and $<$ is a lexicographic order on $\mathcal{A}$ such that $a = \min(\mathcal{A})$.

Definition (Glen, 2006)

An infinite word $\mathbf{t}$ on $\mathcal{A}$ is said to be *fine* if there exists an infinite word $\mathbf{s}$ such that $\min(\mathbf{t}) = a\mathbf{s}$ for any acceptable pair $(a, <)$.

Proposition (Glen, 2006)

An infinite word $\mathbf{t}$ is fine if and only if $\mathbf{t}$ is either a strict episturmian word, or a strict episkew word.

Generalized fine words

Definition

An *acceptable pair* is a pair $(a, <)$ where a is a letter and $<$ is a lexicographic order on $\mathcal{A}$ such that $a = \min(\mathcal{A})$.

Definition (Glen, 2006)

An infinite word $\mathbf{t}$ on $\mathcal{A}$ is said to be *fine* if there exists an infinite word $\mathbf{s}$ such that $\min(\mathbf{t}) = a\mathbf{s}$ for any acceptable pair $(a, <)$.

Proposition (Glen, 2006)

An infinite word $\mathbf{t}$ is fine if and only if $\mathbf{t}$ is either a strict episturmian word, or a strict episkew word.

Generalized fine words

Definition

An *acceptable pair* is a pair $(a, <)$ where a is a letter and $<$ is a lexicographic order on $\mathcal{A}$ such that $a = \min(\mathcal{A})$.

Definition (Glen, 2006)

An infinite word $\mathbf{t}$ on $\mathcal{A}$ is said to be *fine* if there exists an infinite word $\mathbf{s}$ such that $\min(\mathbf{t}) = a\mathbf{s}$ for any acceptable pair $(a, <)$.

Proposition (Glen, 2006)

An infinite word $\mathbf{t}$ *is fine if and only if* $\mathbf{t}$ *is either a strict episturmian word, or a strict episkew word.*

Outline

Terminology

Notation

- Let w be a finite or infinite word on $\mathcal{A}$.
- $\min(w|k)$ denotes the lexicographically smallest factor of w of length k for the given order (where $|w| \geq k$ for w finite).

Definition

- For a finite word $w \in \mathcal{A}^+$ and a given order, $\min(w)$ will denote $\min(w|k)$ where k is maximal such that all $\min(w|j)$, $j = 1, 2, \ldots, k$, are prefixes of $\min(w|k)$.
- In the case $\mathcal{A} = \{a, b\}$, $\max(w)$ is defined similarly.

Terminology

Notation

- Let w be a finite or infinite word on $\mathcal{A}$.
- $\min(w|k)$ denotes the lexicographically smallest factor of w of length k for the given order (where $|w| \geq k$ for w finite).

Definition

- For a finite word $w \in \mathcal{A}^+$ and a given order, $\min(w)$ will denote $\min(w|k)$ where k is maximal such that all $\min(w|j)$, $j = 1, 2, \ldots, k$, are prefixes of $\min(w|k)$.
- In the case $\mathcal{A} = \{a, b\}$, $\max(w)$ is defined similarly.

Example

Suppose $w = baabacababac$.

For the orders $b < a < c$ and $b < c < a$ on the 3-letter alphabet $\{a, b, c\}$:

$$\begin{aligned}
\min(w|1) &= b \\
\min(w|2) &= ba \\
\min(w|3) &= bab \\
\min(w|4) &= baba \\
\min(w|5) &= babac = \min(w)
\end{aligned}$$

Note: $\min(w)$ is a suffix of w, which is true in general.

Characterizations

Notation

v_p: prefix of length p of a given finite or infinite word v.

Theorem

A finite word w on $\mathcal{A}$ is episturmian if and only if there exists a finite word u such that, for any acceptable pair $(a, <)$, we have

$$au_{|m|-1} \leq m \qquad (1)$$

where $m = \min(w)$ for the considered order.

Characterizations

Notation

v_p: prefix of length p of a given finite or infinite word v.

Theorem

A finite word w on $\mathcal{A}$ is episturmian if and only if there exists a finite word u such that, for any acceptable pair $(a, <)$, we have

$$au_{|m|-1} \leq m \tag{1}$$

where $m = \min(w)$ for the considered order.

Example

Recall $w = baabacababac$.

- For the different orders on $\{a, b, c\}$:
  - $a < b < c$ or $a < c < b$: $\min(w) = aabacababac$;
  - $b < a < c$ or $b < c < a$: $\min(w) = babac$;
  - $c < a < b$ or $c < b < a$: $\min(w) = cababac$.
- $u = abacaaaaa$ satisfies (1) $\Rightarrow$ w is finite episturmian.

Characterizations (cont.)

- A new characterization of finite Sturmian words (i.e., finite balanced words):

Corollary

A finite word w on $\mathcal{A} = \{a, b\}$, $a < b$, is not Sturmian (i.e., not balanced) if and only if there exists a finite word u such that aua is a prefix of $\min(w)$ and bub is a prefix of $\max(w)$.

Examples

Example (1)

For $w = ababaabaabab$:

- $\min(w) = a\underline{abaabab}$, $\max(w) = b\underline{abaaba}abab$.
- $abaaba$ is the longest common prefix of $a^{-1}\min(w)$ and $b^{-1}\max(w)$.
- $abaaba$ is followed by b in $\min(w)$ and a in $\max(w)$.
- Thus w is Sturmian.

Example (2)

For $w = aabababaabaab$:

- $\min(w) = aabaab$, $\max(w) = bababaabaab$.
- $\min(w) = \underline{aua}b$ and $\max(w) = \underline{bub}aabaab$ where $u = aba$.
- Thus w is **not** Sturmian.

Examples

Example (1)

For $w = ababaabaabab$:

- $\min(w) = a\underline{abaaba}b$, $\max(w) = b\underline{abaaba}abab$.
- $abaaba$ is the longest common prefix of $a^{-1}\min(w)$ and $b^{-1}\max(w)$.
- $abaaba$ is followed by b in $\min(w)$ and a in $\max(w)$.
- Thus w is Sturmian.

Example (2)

For $w = aababababaabaab$:

- $\min(w) = aabaab$, $\max(w) = bababaabaab$.
- $\min(w) = \underline{au}ab$ and $\max(w) = \underline{bu}baabaab$ where $u = aba$.
- Thus w is **not** Sturmian.

Outline

- A characterization of episturmian words in a *wide sense* (recurrent, episkew):

Corollary

An infinite word $\mathbf{t}$ *on* $\mathcal{A}$ *is episturmian in the wide sense if and only if there exists an infinite word* $\mathbf{u}$ *such that* $a\mathbf{u} \leq \min(\mathbf{t})$ *for any acceptable pair* $(a, <)$.

- A characterization of balanced infinite words on a 2-letter alphabet (i.e., Sturmian and skew words):

Corollary

An infinite word $\mathbf{t}$ *on* $\{a, b\}$, $a < b$, *is balanced (i.e., Sturmian or skew) if and only if there exists an infinite word* $\mathbf{u}$ *such that*

$$a\mathbf{u} \leq \min(\mathbf{t}) \leq \max(\mathbf{t}) \leq b\mathbf{u}.$$

- A characterization of episturmian words in a *wide sense* (recurrent, episkew):

Corollary

An infinite word $\mathbf{t}$ *on* $\mathcal{A}$ *is episturmian in the wide sense if and only if there exists an infinite word* $\mathbf{u}$ *such that* $a\mathbf{u} \leq \min(\mathbf{t})$ *for any acceptable pair* $(a, <)$.

- A characterization of balanced infinite words on a 2-letter alphabet (i.e., Sturmian and skew words):

Corollary

An infinite word $\mathbf{t}$ *on* $\{a, b\}$, $a < b$, *is balanced (i.e., Sturmian or skew) if and only if there exists an infinite word* $\mathbf{u}$ *such that*

$$a\mathbf{u} \leq \min(\mathbf{t}) \leq \max(\mathbf{t}) \leq b\mathbf{u}.$$

- A characterization of episturmian words in a *wide sense* (recurrent, episkew):

Corollary

An infinite word $\mathbf{t}$ *on* $\mathcal{A}$ *is episturmian in the wide sense if and only if there exists an infinite word* $\mathbf{u}$ *such that* $a\mathbf{u} \leq \min(\mathbf{t})$ *for any acceptable pair* $(a, <)$*.*

- A characterization of balanced infinite words on a 2-letter alphabet (i.e., Sturmian and skew words):

Corollary

An infinite word $\mathbf{t}$ *on* $\{a, b\}$*,* $a < b$*, is balanced (i.e., Sturmian or skew) if and only if there exists an infinite word* $\mathbf{u}$ *such that*

$$a\mathbf{u} \leq \min(\mathbf{t}) \leq \max(\mathbf{t}) \leq b\mathbf{u}.$$